

## Weekly Report

### Done

1. 阅读迁移学习手册. 仅从这些方法角度而言我还没有看出一些直接能用的东西, 因为这些算法仍然十分和场景有关.
  - 暂时能直接想到的是可能可以将图像上一些模型迁移到关会华那边的图标特征提取中去, 但是这个也不算是他们当前工作重点, 而且对于他们这种企业需要实际做产品的项目来说, 用黑盒的 NN 有一定风险在里面.
2. 巫老师认为我当前可以做一些对于可视化数据\任务\系统做一些 code 的基础工作. 简单点说, 希望将他们形式化, 以便于后面能运用各种学习.
  - a. 从近的说, 我们如果能用合适的方式将这三者形式化, 然后建立或者训练一些彼此的关系, 便可以做推荐. 这里的推荐比如说, 公司给你数据, 你也不知道要做啥, 那我怎样根据数据提供一些可视化可以做的任务, 以方便我和公司来进行交流和迭代? 有了任务, 我又可以怎样的设计一些原型的可视化系统, 使得系统的可视化和交互两部分能够和我的任务对应起来? 我只要做原型系统, 后续的不用管.
  - b. 这是一个比较大但也比较痛点的问题. 以我本周看的一些文献而言, (1) 做可视化推荐的也都不少了, 但是多的还是简单图表, 涉及不到复杂可视化; (2) 而且, 一般也不会涉及到交互. (3) 当然整个这一个流程没有现成的
  - c. 光从数据或者任务而言, 已有有不少文章做了一个形式化的设计; 数据方面可能不需要的可视化领域找就有很多, 当然可视化方面也有从数据的不同角度来视图做推荐的, 比如数据类型\数据分布等等. 任务方面就有像 Tamara 一些综述以及 A Design Space of Visualization Tasks 这样的文章. 这几篇都略有印象, 仍在研读之中. 需要注意的是, 这里的任务需要有一定的抽象程度.
  - d. 对于可视化系统的形式化, (1) 直接能想到的是拆解, 这方面今年 infovis 有一篇工作 NarVis 有类似的操作, 但是由于其面向对象是教学性质的, 我们仍可以有一些不同的思路. (2) 如果能用 declarative 的语言来形式化, 那是比较好的; 但是仍然要考虑其他一些形式. (2.1) 对于之前提到的 draco, 仍然不是觉得很能用, 也许在后面真正到了要做推荐这一步时, 可以效仿一下这种 constraint 的思路; (3) 此外一定要注意, 对于可视化的形式化, 我需要抽象到哪一层的问题, 过于细和过于粗都对后面的建立规则是一个威胁.
  - e. NarVis 可以学习以及的地方
    - i. Hierarchical 的拆解可视化
    - ii. 本文仅限于一个视图
    - iii. Visual units 之间的关联
    - iv. 基于 salience 的介绍顺序排序以及 Progressive introduction of visual channels. 尽管从形式上想, 本文是教育性质的可视化拆解工具, 似乎与我们要形式化关系不大, 但是仍然有可能这样的拆解方式本身就有一些内在优点在, 所以还要继续思考一下
  - f. 对于整个形式化的工作, 必然需要大量劳动力; 所以另一个思路是用 NLP 来提取文本信息(数据描述, 任务), 再用视觉方法提取可视化结构, 来帮助偷懒.
3. 对于 fusion, 我和许多同学有聊了一下, 但由于时间仓促, 还没有特别深入思考. 有以下两个角度.

- a. 光从散点图角度来说(切入点简单), 散点图也有各种简化且提升用户感知的方式, 毕竟散点图点一多很多就看不清了. 结合海东师兄当年那篇蓝噪声采用, 其实里面有各种技术, 包括可以从散点图到热力图无极变换的 splatter plot 等(我觉得我们可以简单的说热力图是一种数据融合后再绘制而散点图是直接画出来的). 如果能有一定任务来做这些图的 user study; 或者再找一个例子, 找到其与散点图的共通性, 相当于可以 derive 一些数据融合与可视化融合的准则
  - b. 另外回到 NodeTrix, 那天粗略寻找一下也没有比较好的 nodetrix 如何控制何时 node 何时画 matrix 的文章. 如果网络一大, 点本身位置, 到底这一个 community 用点还是矩阵来画, 矩阵内部如何更好的排列可能都是一个问题, 还可能还有其他问题. 仅以今年郭博文章那个网络为例, 其网络之中仍有一些文章没有探索的区域, 其 layout 十分 clutter, 但这还只是整个网络的一部分.
4. 安排小顾学习 tensorflow 一些入门知识. 本周搭建环境, 学习前面 shallow model 在 tensorflow 中的了解与实现, 以及对神经网络基本知识的学习.

## To Do

1. 看一些文章, 主要是这周残余的与任务分类和可视化推荐的文章
2. 讨论上面的 2, 3. 对于 2 整个问题需要一个可行性的认识以及一个初步的思路线(尤其是可视化部分), 对于 3 的两个角度会分别请教老师和同学.

## 论文阅读

*Beyond heuristics: learning visualization design* Arxiv 综述性质短文章, 概述现有可视化推荐系统, 以及方法论, 以及未来可能的自动化方法. 其提到的 markov logic network 我还在看, 也是一种基于网络的推断方法. 提出了一些开放性问题

*VizDeck: Self-Organizing Dashboards for Visual Analytics* 一个可视化推荐系统, 希望根据数据分布和用户投票(喜好)来学习一个推荐系统. 里面有一个类似 provenance view 的小视图以及用户协同的功能, 很有意思, 因为可视化构建仍然是一个过程, 一个有偏好的过程, 在于他人分享时需要展现其过程才能更好体现其背后思路. 这一点在很多 provenance vis 可视化中都有体现.

*NarVis: Authoring Narrative Slideshows for Introducing Data Visualization Designs* 上面已经提到他们的点

其余如迁移学习手册, 西瓜书上的贝叶斯网络, 以及各种 voyager, voyager2 和其余一些相关资料的浏览.

## 工作时间

平时 9.5, 周末一共 12 小时. 总共 60.

## 个人规划

性质	Ddl	进展	目标
中长期	2019 年中	统计: 假设检验导论	统计 优化 算法

短期	11.17	两块: 散点图和图上的问题, 详见上面	Fusion 相关
短期	18 年底	基本的形式化工作	XX 学习对可视化支持
长期			博士毕业论文中的研究方向